A Survey and a Credit Card Fraud Detection and Prevention Model using the Decision Tree Algorithm

Abdulaziz Saleh Alraddadi
College of Computer Science and Engineering at Yanbu, Taibah University, Saudi Arabia
alraddadi1@yahoo.com (corresponding author)

Received: 17 June 2023 | Revised: 27 June 2023 | Accepted: 1 July 2023
Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: https://doi.org/10.48084/etasr.6128

ABSTRACT

Today, many people prefer online payment methods due to the rapid growth in cashless electronic transactions. Credit and debit cards are the most popular electronic payment methods but are prone to fraud due to the nature of their use and the tendency of fraudsters to access their details. This study proposes a theoretical credit fraud detection and prevention model using a Decision Tree Algorithm (DCA). Moreover, a survey questionnaire was used to investigate students’ perceptions of credit card fraud incidents. Data were collected from 102 students from different universities and countries around the world. The results showed that 95.9% of the respondents knew how credit/debit card fraud occurs, while 4.1% of them did not. Finally, 81.6% expressed their willingness to use a tool based on the proposed model to prevent or detect credit/debit card fraud incidents.

Keywords-online payment; credit card fraud; Decision Tree Algorithm (DCA); survey

I. INTRODUCTION

Cybercrime is one of the most serious threats cyberspace users and the global economy face. According to European and international organizations, it is considered one of the most significant security challenges of the 21st century [1]. Numerous studies have attempted to investigate or prevent cybercrime methods. Such studies contain database schemes [2], automation [3], avoiding cyberbullying [4], wireless networks [5-6], cloud safety [7-8], smart IoT [9-10], drone domain and mobile areas [11], and the health field [12-20]. Companies and organizations monitor consumers’ or users’ spending patterns to identify and prevent fraudulent activities. Credit cards are becoming increasingly popular as a payment method in both online and conventional transactions. Fraud detection involves the identification and capture of fraudulent activities and events. In recent decades, many modern methods have been developed to prevent frauds related to credit card transactions. Several approaches can be used to detect and prevent fraudulent transactions, including machine learning, data mining, sequence alignment, fuzzy reasoning, genetic programming, and fuzzy reasoning [21-22]. Fraud occurs when goods, services, or money are acquired illegally. There are many situations in which frauds occur when there is criminal intent present, but it is often obscure. In addition to credit cards, there are many other fraud targets. Credit card fraud is a general term for theft, fraud, and other similar payment methods used fraudulently to fund a transaction. Credit card fraud is becoming a growing problem. As it can be challenging to identify credit card fraud using standard procedures, the development of prevention models is now crucial for academic and commercial institutions. Furthermore, with the development of technology over the past few decades, fraud methods have changed dramatically. Credit card fraud is one of the greatest obstacles for business and commercial enterprises. Credit card fraud is the use of another person's credit card for private reasons without the person who owns the card or the card issuer being aware of it [23-24].

Credit card fraud can be eliminated, and financial risks can be reduced using many systems, models, procedures, and preventive measures. Although the number of people using credit cards is quickly increasing, very few feel secure and confident in using them for regular purchases [23, 25]. Credit card fraud is expanding around the world along with advancements in communication channels and information technology, leading to significant losses. This study investigated the extent to which students have been victims of credit card fraud and focused on a model to detect and prevent credit card fraud on campuses. To address this objective, four research questions (RQ) were formulated as follows:

- **RQ1:** Have you ever heard of or been involved in credit card fraud before?
- **RQ2:** How do you think it does happen?
- **RQ3:** How do you think it can be mitigated?
- **RQ4:** Would you use a solution based on the proposed model that prevents credit card fraud?

The main contribution of this study is the proposal of a theoretical framework for a model to prevent credit card fraud.
Moreover, this study investigated students' perceptions of how credit card fraud occurs and can be prevented or detected using a survey questionnaire. Finally, it was determined that the students would be willing to use such a system.

II. RELATED WORKS

In [26], an algorithmic framework was developed using a decision tree and a combination of Luhn's and Hunt's algorithms to prevent credit card fraud. Frauds were detected using Luhn's algorithms in incoming transactions, which were assessed based on address mismatches and degrees of outliers to determine how far they deviate from the typical customer profile. Bayes' theorem was applied to strengthen or weaken the general belief, and a more sophisticated combinatorial heuristic was applied to combine the probable fraud with the original belief. In [27], three methods were proposed, using machine learning algorithms to identify fraudulent transactions. Many metrics can be used to assess how well a classifier or predictor performs, such as the vector machine, the random forest, and decision trees. It is important to note that these measures depended on prevalence [28]. In [29], a comparative study was conducted among the different methods of detecting credit card fraud. In [30], three security levels were introduced in the concept of credit card fraud detection based on a Hidden Markov Model (HMM). In [31], HMMs were used to detect fraudulent card payments in two stages using dynamic random forest and k-Nearest-Neighbor algorithms. In [32], advanced data mining techniques were investigated to combine with a neural network based on real-time credit card data to offer even stronger predictive capabilities. In [33], two kinds of random forest methods were used to train features that represent normal or abnormal transactional behavior. In [34], supervised algorithms were used for the same purpose, including Deep Learning, Logistic Regression, Nave Bayesian, Support Vector Machines (SVM), Neural Networks, Artificial Immune Systems, K-Nearest Neighbor, Decision Trees, Data Mining, and Fuzzy logic-based systems. In [27, 35] prediction, clustering, and outlier identification were compared in machine learning techniques. Random forests were used to train a classifier to recognize the behavioral characteristics of credit card transactions. The following methods were used to train the genuine and fraudulent "behavior feature": CART-based random forest and random forest based on random trees. Transactions were grouped into relevant sliding window groups, and various window features were retrieved to discover consumer behavioral patterns. There are features such as the maximum or minimum amount of a transaction, average amounts in the windows, and even the time elapsed [36].

Systems generate fraud scores for specific transactions using a variety of instructions and algorithms. In [37], deep neural network techniques were proposed for fraud detection. The preparation approach is a significant setback in addressing data skew issues in a dataset. There are numerous methods to assess whether a transaction is legitimate or fraudulent [24]. The drawback of supervised learning is that it depends on people to optimize its settings. On the contrary, decision trees can be built more quickly than other techniques and do not require the user to set any parameters [26]. Although internet payments are more convenient, adequate, and easy to use, losses related to electronic commerce should not be ignored. There are many security frameworks for businesses and banks, but fraudsters change their subtle strategies over time to bypass them. Therefore, it is essential to improve methods for detection and prevention [28]. Understanding the way a fraud is carried out is essential for effective fraud prevention. A credit card fraud detection tool should depend on the fraud technique itself [38-39]. In [40-42], machine learning-based methods were used to detect fraud in financial transactions and analyze research gaps to uncover research trends in the area. In [46], some unsupervised credit card fraud detection methods were proposed with the aid of behavioral outlier identification methods. Most likely, fraud incidents will be discovered as anomalous spending patterns and transaction frequency.

<table>
<thead>
<tr>
<th>Ref</th>
<th>Focus</th>
</tr>
</thead>
<tbody>
<tr>
<td>(26)</td>
<td>Credit cards were validated using the credit card number (input). Incoming transactions were assessed based on address mismatches and degrees of outliers to determine how far they deviate from the typical customer profile.</td>
</tr>
<tr>
<td>(27)</td>
<td>Many metrics were used to evaluate the performance of a classifier or predictor.</td>
</tr>
<tr>
<td>(30)</td>
<td>Three levels of security were introduced using a HMM.</td>
</tr>
<tr>
<td>(31)</td>
<td>Dynamic random forest algorithms and k-nearest neighbor algorithms were used to detect credit card fraud in two stages.</td>
</tr>
<tr>
<td>(32)</td>
<td>The possibility of improving prediction by combining advanced data mining techniques with a neural network based on real-time credit card data was investigated.</td>
</tr>
<tr>
<td>(33)</td>
<td>Two types of random forest methods were used. Training features represent normal and abnormal transaction behavior.</td>
</tr>
</tbody>
</table>

III. THE PROPOSED MODEL

This study used design science research [43] to propose a credit fraud detection and prevention model, using a Decision Tree Algorithm (DCA). The development process consisted of three main stages, as shown in Figure 1.

![Research methodology](image)

Fig. 1. Research methodology.
These stages were: collecting data, filtering the collected data, and selecting detection models for implementation. In the first stage, common search engines were used to retrieve relevant articles using the keywords "Fraud Prevention" and "Fraud Detection" for the period 2014-2023, and Kaggle was used to retrieve several credit card transaction datasets with labels for fraudulent credit card transactions. Only journal and conference papers written in English were collected. The datasets had several credit card transactions, with an additional binary field called that indicated whether a transaction was fraudulent (1) or not. The datasets also included transaction features, such as amount, time, frequency, geographic location, merchant information, cardholder information, transaction type, and device type and information. In the second stage, the collected data were filtered. A DCA was used in the design of the credit fraud detection and prevention model based on common features, as shown in Figure 2.

The model used supervised learning, specifically DCA. DCA tree-structured classification starts with a root node and divides the dataset's features into internal nodes that represent dataset features, decision instructions, and results [44]. A decision tree simply poses a query and divides it into subtrees depending on the response. Although DCA can resolve classification and regression problems, it is mostly applied to the former. The method starts its search at the top of the tree to locate the dataset classes. To reach the next node, it calculates the branch based on the relations and matches the base trait with the record attribute [45]. As shown in Figure 2, the proposed model was designed to work on the user's device as a client-side implementation to forecast the possibility of fraud. Based on the forecast made by the model, the user gets real-time feedback regarding the transaction risk level. If the model detects a high likelihood of fraud, the user may receive a warning. The model should be compatible with Android, iOS, and PC operating systems. Credit or debit card transactions were assumed to take place over the Internet, at a sales point, or in a mobile application. At the initiation of the transaction, the card status is checked, and the transaction proceeds if the card is valid. The model uses DCA, the expected score is generated, the investigation is carried out, and the appropriate decision is taken. If the score is appropriate, the card status is further checked. If a fraudulent trace is noticed, it automatically rejects the transaction, otherwise it is executed and outputs successfully.

The second phase used a questionnaire that was distributed through the WhatsApp platform, targeting a population of 450 students. The adopted questionnaire design was vetted by an expert and 102 responses were recovered within 48 hours. The purpose of the survey was to obtain responses that will provide answers to research questions. The proposed model was discussed with the respondents before sending the questionnaires.

IV. SURVEY

The survey was carried out among computer science students. Figure 3 shows the ages of the respondents, where 79.6% were 21-30 years old, and 20.4% were 11-20 years old, and Figure 4 shows the degree levels of the participant students.

Fig. 2. The proposed credit card fraud detection and prevention model.

Fig. 3. Ages of respondents.

Fig. 4. Respondents' degree levels.
Developing a system that uses fingerprint before card details are requested to identify the user of the mobile application or on-site.

- Other personal details of the user should be required in online transactions rather than the serial number of the card and CVV.

- Cyber-security awareness should be consolidated for everyone, including students and business owners.

- Individuals should be careful when using credit cards online, especially when certain malware can be easily installed unknowingly, such as keyloggers or spyware.

- The ability of access to customers' details after a withdrawal at POS centers should be disabled.

- Educating people about not disclosing their details and avoiding visiting sites that are not secure.

V. CONCLUSION

Credit card fraud is an issue that has received a lot of attention during the recent years. This study presented a theoretical model to minimize the challenges associated with credit card fraud and surveyed the perspectives of university students. This study also provides insight into the willingness of the respondents to use a method that will provide such prevention measures. As a result of the study, information was obtained from selected university groups of students and the results indicate that students are very positive about using a method based on the proposed model. Furthermore, the study offers suggestions that could help prevent users from being scammed by debit/credit card fraud. In the future, the proposed model could be developed and tested in a real-world scenario.

REFERENCES


